

FIGURE 1A

Nucleotide Sequence of Human ABCG4 Transporter Gene

Sequence Range: 1 to 3455

GCCACCATGG CGGAGAAGGC GCTGGAGGCC GTGGGCTGTG GACTAGGGCC GGGGGCTGTG 60
GCCATGGCCG TGACGCTGGA GGACGGGGCG GAACCCCTG TGCTGACCAC GCACCTGAAG 120
AAGGTGGAGA ACCACATCAC TGAAGCCCAG CGCTTCTCCC ACCTGCCCAA GCGCTCAGCC 180
GTGGACATCG AGTTCGTGGA GCTGTCTTAT TCCGTGCGGG AGGGGCCCTG CTGGCGCAAA 240
AGGGTTTATA AGACCCTTCT CAAGTGCCTC TCAGGTAAAT TCTGCCGCCG GGAGCTGATT 300
GGCATCATGG GCCCTCAGG GGCTGGCAAG TCTACATTCA TGAACATCTT GGCAGGATAC 360
AGGGAGTCTG GAATGAAGGG GCAGATCCTG GTTAATGGAA GGCCACGGGA GCTGAGGACC 420
TTCCGCAAGA TGTCTGCTA CATCATGCAA GATGACATGC TGCTGCCGCA CCTCACGGTG 480
TTGGAAGCCA TGATGGTCTC TGCTAACCTG AATCTTACTG AGAATCCCGA TGTGAAAAAC 540
GATCTCGTGA CAGAGATCCT GACGGCACTG GGCCTGATGT CGTGCTCCCA CACGAGGACA 600
GCCCTGCTCT CTGGCGGGCA GAGGAAGCGT CTGGCCATCG CCCTGGAGCT GGTCAACAAC 660
CCGCCTGTCA TGTCTTTGA TGAGCCCACC AGTGGTCTGG ATAGCGCCTC TTGTTTCCAA 720
GTGGTGTCCC TCATGAAGTC CCTGGCACAG GGGGGCCGTA CCATCATCTG CACCATCCAC 780
CAGCCCAGTG CCAAGCTCTT TGAGATGTTT GACAAGCTCT ACATCCTGAG CCAGGGTCAG 840
TGCATCTTCA AAGGAGTGGT CACCAACCTG ATCCCCTATC TAAAGGGACT CGGCTTGCA 900
TGCCCCACCT ACCACAACCC GGCTGACTTC ATCATCGAGG TGGCCTCTGG CGAGTATGGA 960
GACCTGAACC CCATGTTGTT CAGGGCTGTG CAGAATGGGC TGTGCGCTAT GGCTGAGAAG 1020
AAGAGCAGCC CTGAGAAGAA CGAGTCCCT GCCCATGCC CTCCTTGTC TCCGGAAGTG 1080
GATCCATTG AAAGCCACAC CTTTGCCACC AGCACCTCA CACAGTCTG CATCCTCTTC 1140
AAGAGGACCT TCCTGTCCAT CCTCAGGGAC ACGGTCCTGA CCCACCTACG GTTCATGTCC 1200
CACGTGGTTA TTGGCGTGCT CATCGGCCTC CTCTACCTGC ATATTGGCGA CGATGCCAGC 1260
AAGGTCTTCA ACAACACCGG CTGCCTCTTC TTCTCCATGC TGTTCTCAT GTTCGCCGCC 1320
CTCATGCCAA CTGTGCTCAC CTTCCTCTTA GAGATGGCGG TCTTCATGAG GGAGCACCTC 1380
AACTACTGGT ACAGCCTCAA AGCGTATTAC CTGGCCAAGA CCATGGCTGA CGTGCCCTTT 1440
CAGGTGGTGT GTCCGGTGGT CTACTGCAGC ATTGTGTACT GGATGACGGG CCAGCCCGCT 1500
GAGACCAGCC GCTTCCTGCT CTTCTCAGCC CTGGCCACCG CCACCGCCTT GGTGGCCCAA 1560
TCTTTGGGGC TGCTGATCGG AGCTGCTTCC AACTCCCTAC AGGTGGCCAC TTTTGTGGGC 1620
CCAGTTACCG CCATCCTGT CCTCTGTTC TCCGGCTTCT TTGTGAGCT CAAGACCATC 1680
CCCCTTACC TGCAATGGAG CTCCTATCTC TCCTATGTCA GGTATGGCTT TGAGGGTGTG 1740
ATCCTGACGA TCTATGGCAT GGAGCGAGGA GACCTGACAT GTTTAGAGGA ACGCTGCCCG 1800
TTCCGGGAGC CACAGAGCAT CCTCCGAGC CTGGATGTGG AGGATGCCAA GCTCTACATG 1860
GACTTCCTGG TCTTGGGCAT CTTCTCCTA GCCCTGCGGC TGCTGGCCTA CCTTGTGCTG 1920
CGTTACCGGG TCAAGTCAGA GAGATAGAGG CTTGCCCCAG CCTGTACCCC AGCCCCTGCA 1980
GCAGGAAGCC CCCAGTCCA GCCCTTTGGG ACTGTTTTAA CCTTATAGAC TTGGGCACTG 2040
GTTCTTGGCG GGGCTATCCT CTCCTCCCTT GGCTCCTCCA CAGGCTGGCT GTCGGACTGC 2100
GCTCCCAGCC TGGGCTCTGG GAGTGGGGGC TCCAGCCCTC CCCACTATGC CCAGGAGTCT 2160
TCCCAAGTTG ATGCGGTTTG TAGCTTCCTC CTTACTCTCT CCAACACCTG CATGCAAGA 2220
CTACTGGGAG GCTGCTGCCT CTTCTGCC CATGGCACCC TCCTCTGCTG TCTGCCTGGG 2280
AGCCCTAGGC TCTCTAGGGC CCCACTTACA ACTGACCAA GTGGCCCCCT CTGGGGGTCC 2340
CCACCACACA AGTGTGTGTA AACTGGGCTG CTATAAGGTT GGAGTTCCAG GGCTGGGCCC 2400
TGGTGGAGTC CACTGGAAGT CCCATTATGG ATGTTGAAAT GGACAGGGAA GGACTCTGGA 2460
AGTCTCTTCC TCCTCCTCCT CTTCTCTCCA CCCCTAGACC CTGGCTGACT TGGACAATCT 2520
GCCAGGACAG AAGCTGGGTT TTCTGTCTAG GTCACCACTC CCAATCCTGG GGATTGGAGA 2580
GGCTGGGGC TGTGGGATGC CCCATCCCC TCCCCATCAC CTTTGGTGGG GGCAGGGCCT 2640
GGTGGCACCT GTGCAATAAT GTCTGTGTTT CTCTCCCACC TGCCACTGGA ACTGGAGAAT 2700
GCACTTTATT CTGGGCGGGG GGTGAGTGGG GGAAGACCCA ACCCTCCTTT CTCGCTGCCC 2760
CTAACGCATG CACGGTCTCG TGATGCTCCC TCCCTCTCCG GAGTGACAGG CACATACATG 2820
AGAACAGGCC ATCTCAGCCC TACACACTTG CCATCCCCTA CAGCACAGAG GAAGAGTGAT 2880
GGTGGCATGC TGGTGGTGGC GGGTGTGCTGG GGGAGGACAG TGCCAACCTC CTCCTGGGGA 2940

20040505: 5406001

FIGURE 1B

TCCCATGTTG GAGACTCTAA GGATAAGGCT GGTGCTGCCC AGGGTGTCTA CAGGAACTGC 3000
AGGTGTCTAC CCCCAAGTCT TCCCTCCTCC CAAGCCAGGG GTGGCACAGG GCACTAGATC 3060
CCTGGAGTTC AGGAACCAAC ACAAGCACAA CCACGGGCAT AAGTTGGCCT TGGCCACTGC 3120
CACCCACGGC CCTCCTTTTG TGCTCCATGC TGGCATCTTC ACTCCCCTAC CCCTTCCCCA 3180
GCCACTGCTG CTCATTCAAA CTTCTGTCCA TGTCCCTCCA CTGTTCTTAT CAGCAGGTGG 3240
CCCCTGGGCA TCAGAACAGC CTGCCCTGGG CACCAGGTGG CAGACACACT CAGAGCATGT 3300
CTGGCTTTCC TGGTGGGTCC AGGCTCATTC TGCTTCTGAT TTCCCCTCCC CCAGGGCTCA 3360
TTTCCCCCT TTTTCCTGTA CACATCCCTG TCTACCTCCT CTCACCCTGC CACAGATTCT 3420
TCCTATCACA CAGGGATGCC AGTTGTATTT GTGGG 3455

2007-05-03 14:00:00

FIGURE 2A

Coding Sequence of Human ABCG4 Transporter Gene

Sequence Range: 1 to 3455

gcc acc atg gcg gag aag gcg ctg gag gcc gtg ggc tgt gga cta ggg ccg ggg gct gtg 60
Met Ala Glu Lys Ala Leu Glu Ala Val Gly Cys Gly Leu Gly Pro Gly Ala Val

gcc atg gcc gtg acg ctg gag gac ggg gcg gaa ccc cct gtg ctg acc acg cac ctg aag
120

Ala Met Ala Val Thr Leu Glu Asp Gly Ala Glu Pro Pro Val Leu Thr Thr His Leu Lys

aag gtg gag aac cac atc act gaa gcc cag cgc ttc tcc cac ctg ccc aag cgc tca gcc
180

Lys Val Glu Asn His Ile Thr Glu Ala Gln Arg Phe Ser His Leu Pro Lys Arg Ser Ala

gtg gac atc gag ttc gtg gag ctg tcc tat tcc gtg cgg gag ggg ccc tgc tgg cgc aaa
240

Val Asp Ile Glu Phe Val Glu Leu Ser Tyr Ser Val Arg Glu Gly Pro Cys Trp Arg Lys

agg ggt tat aag acc ctt ctc aag tgc ctc tca ggt aaa ttc tgc cgc cgg gag ctg att
300

Arg Gly Tyr Lys Thr Leu Leu Lys Cys Leu Ser Gly Lys Phe Cys Arg Arg Glu Leu Ile

ggc atc atg gcc ccc tca ggg gct gcc aag tct aca ttc atg aac atc ttg gca gga tac
360

Gly Ile Met Gly Pro Ser Gly Ala Gly Lys Ser Thr Phe Met Asn Ile Leu Ala Gly Tyr

agg gag tct gga atg aag ggg cag atc ctg gtt aat gga agg cca cgg gag ctg agg acc
420

Arg Glu Ser Gly Met Lys Gly Gln Ile Leu Val Asn Gly Arg Pro Arg Glu Leu Arg Thr

ttc cgc aag atg tcc tgc tac atc atg caa gat gac atg ctg ctg ccg cac ctc acg gtg
480

Phe Arg Lys Met Ser Cys Tyr Ile Met Gln Asp Asp Met Leu Leu Pro His Leu Thr Val

2000-01-01

ttg gaa gcc atg atg gtc tct gct aac ctg aat ctt act gag aat ccc gat gtg aaa aac
540
Leu Glu Ala Met Met Val Ser Ala Asn Leu Asn Leu Thr Glu Asn Pro Asp Val Lys Asn
gat ctg gtg aca gag atc ctg acg gca ctg ggc ctg atg tcg tgc tcc cac acg agg aca
600
Asp Leu Val Thr Glu Ile Leu Thr Ala Leu Gly Leu Met Ser Cys Ser His Thr Arg Thr
gcc ctg ctg tct ggc ggg cag agg aag cgt ctg gcc atc gcc ctg gag ctg gtc aac aac
660
Ala Leu Leu Ser Gly Gly Gln Arg Lys Arg Leu Ala Ile Ala Leu Glu Leu Val Asn Asn
ccg cct gtc atg ttc ttt gat gag ccc acc agt ggt ctg gat agc gcc tct tgt ttc caa
720
Pro Pro Val Met Phe Phe Asp Glu Pro Thr Ser Gly Leu Asp Ser Ala Ser Cys Phe Gln
gtg gtg tcc ctg atg aag tcc ctg gca cag ggg ggc cgt acc atc atc tgc acc atc cac
780
Val Val Ser Leu Met Lys Ser Leu Ala Gln Gly Gly Arg Thr Ile Ile Cys Thr Ile His
cag ccc agt gcc aag ctg ttt gag atg ttt gac aag ctg tac atc ctg agc cag ggt cag
840
Gln Pro Ser Ala Lys Leu Phe Glu Met Phe Asp Lys Leu Tyr Ile Leu Ser Gln Gly Gln
tgc atc ttc aaa gga gtg gtc acc aac ctg atc ccc tat cta aag gga ctg ggc ttg cat
900
Cys Ile Phe Lys Gly Val Val Thr Asn Leu Ile Pro Tyr Leu Lys Gly Leu Gly Leu His
tgc ccc acc tac cac aac ccg gct gac ttc atc atc gag gtg gcc tct ggc gag tat gga
960
Cys Pro Thr Tyr His Asn Pro Ala Asp Phe Ile Ile Glu Val Ala Ser Gly Glu Tyr Gly
gac ctg aac ccc atg ttg ttc agg gct gtg cag aat ggg ctg tgc gct atg gct gag aag
1020
Asp Leu Asn Pro Met Leu Phe Arg Ala Val Gln Asn Gly Leu Cys Ala Met Ala Glu Lys
aag agc agc cct gag aag aac gag gtc cct gcc cca tgc cct cct tgt cct ccg gaa gtg
1080
Lys Ser Ser Pro Glu Lys Asn Glu Val Pro Ala Pro Cys Pro Pro Cys Pro Pro Glu Val

FIGURE 2C

gat ccc att gaa agc cac acc ttt gcc acc agc acc ctc aca cag ttc tgc atc ctc ttc
1140
Asp Pro Ile Glu Ser His Thr Phe Ala Thr Ser Thr Leu Thr Gln Phe Cys Ile Leu Phe

aag agg acc ttc ctg tcc atc ctc agg gac acg gtc ctg acc cac cta cgg ttc atg tcc
1200
Lys Arg Thr Phe Leu Ser Ile Leu Arg Asp Thr Val Leu Thr His Leu Arg Phe Met Ser

cac gtg gtt att ggc gtg ctc atc ggc ctc ctc tac ctg cat att ggc gac gat gcc agc
1260
His Val Val Ile Gly Val Leu Ile Gly Leu Leu Tyr Leu His Ile Gly Asp Asp Ala Ser

aag gtc ttc aac aac acc ggc tgc ctc ttc ttc tcc atg ctg ttc ctc atg ttc gcc gcc
1320
Lys Val Phe Asn Asn Thr Gly Cys Leu Phe Phe Ser Met Leu Phe Leu Met Phe Ala Ala

ctc atg cca act gtg ctc acc ttc ccc tta gag atg gcg gtc ttc atg agg gag cac ctc
1380
Leu Met Pro Thr Val Leu Thr Phe Pro Leu Glu Met Ala Val Phe Met Arg Glu His Leu

aac tac tgg tac agc ctc aaa gcg tat tac ctg gcc aag acc atg gct gac gtg ccc ttt
1440
Asn Tyr Trp Tyr Ser Leu Lys Ala Tyr Tyr Leu Ala Lys Thr Met Ala Asp Val Pro Phe

cag gtg gtg tgt ccg gtg gtc tac tgc agc att gtg tac tgg atg acg ggc cag ccc gct
1500
Gln Val Val Cys Pro Val Val Tyr Cys Ser Ile Val Tyr Trp Met Thr Gly Gln Pro Ala

gag acc agc cgc ttc ctg ctc ttc tca gcc ctg gcc acc gcc acc gcc ttg gtg gcc caa
1560
Glu Thr Ser Arg Phe Leu Leu Phe Ser Ala Leu Ala Thr Ala Thr Ala Leu Val Ala Gln

tct ttg ggg ctg ctg atc gga gct gct tcc aac tcc cta cag gtg gcc act ttt gtg ggc
1620
Ser Leu Gly Leu Leu Ile Gly Ala Ala Ser Asn Ser Leu Gln Val Ala Thr Phe Val Gly

cca gtt acc gcc atc cct gtc ctc ttg ttc tcc ggc ttc ttt gtc agc ttc aag acc atc
1680

100103.406

Pro Val Thr Ala Ile Pro Val Leu Leu Phe Ser Gly Phe Phe Val Ser Phe Lys Thr Ile
ccc act tac ctg caa tgg agc tcc tat ctc tcc tat gtc agg tat ggc ttt gag ggt gtg
1740
Pro Thr Tyr Leu Gln Trp Ser Ser Tyr Leu Ser Tyr Val Arg Tyr Gly Phe Glu Gly Val
atc ctg acg atc tat ggc atg gag cga gga gac ctg aca tgt tta gag gaa cgc tgc ccg
1800
Ile Leu Thr Ile Tyr Gly Met Glu Arg Gly Asp Leu Thr Cys Leu Glu Glu Arg Cys Pro
ttc cgg gag cca cag agc atc ctc cga gcg ctg gat gtg gag gat gcc aag ctc tac atg
1860
Phe Arg Glu Pro Gln Ser Ile Leu Arg Ala Leu Asp Val Glu Asp Ala Lys Leu Tyr Met
gac ttc ctg gtc ttg ggc atc ttc ttc cta gcc ctg cgg ctg ctg gcc tac ctt gtg ctg
1920
Asp Phe Leu Val Leu Gly Ile Phe Phe Leu Ala Leu Arg Leu Leu Ala Tyr Leu Val Leu
cgt tac cgg gtc aag tca gag aga tag agg ctt gcc cca gcc tgt acc cca gcc cct gca
1980
Arg Tyr Arg Val Lys Ser Glu Arg ***
gca gga agc ccc cag tcc cag ccc ttt ggg act gtt tta acc tta tag act tgg gca ctg
2040
gtt cct ggc ggg gct atc ctc tcc tcc ctt ggc tcc tcc aca ggc tgg ctg tcg gac tgc
2100
gct ccc agc ctg ggc tct ggg agt ggg ggc tcc agc cct ccc cac tat gcc cag gag tct
2160
tcc caa gtt gat gcg gtt tgt agc ttc ctc cct act ctc tcc aac acc tgc atg caa aga
2220
cta ctg gga ggc tgc tgc ctc ctt cct gcc cat ggc acc ctc ctc tgc tgt ctg cct ggg
2280
agc cct agg ctc tct agg gcc cca ctt aca act gac caa agt ggc ccc ctc tgg ggg tcc
2340
cca cca cac aag tgt ttg taa act ggg ctg cta taa ggt tgg agt tcc agg gct ggg ccc
2400

FIGURE 2E

tgg tgg agt cca ctg gaa gtc cca tta tgg atg ttg aaa tgg aca ggg aag gac tct gga
2460
agt ctc ttc ctc ctc ctc ttc tct cca ccc cta gac cct ggc tga ctt gga caa tct
2520
gcc agg aca gaa gct ggg ttt tct gtc tag gtc acc act ccc aat cct ggg gat tgg aga
2580
ggc ctg ggg ctg tgg gat gcc cca tcc ccc tcc cca tca cct ttg gtg ggg gca ggg cct
2640
ggt ggc acc tgt gca ata atg tct gtg ttt ctc tcc cac ctg cca ctg gaa ctg gag aat
2700
gca ctt tat tct ggg cgg ggg gtg agt ggg gga aga ccc aac cct cct ttc tgc ctg ccc
2760
cta acg cat gca cgg tct cgt gat gct ccc tcc ctc tcc gga gtg aca ggc aca tac atg
2820
aga aca ggc cat ctc agc cct aca cac ttg cca tcc cct aca gca cag agg aag agt gat
2880
ggt ggc atg ctg gtg gtg gcg ggt gct ggt ggg agg aca gtg cca acc tcc tcc tgg gga
2940
tcc cat gtt gga gac tct aag gat aag gct ggt gct gcc cag ggt gtc tac agg aac tgc
3000
agg tgt cta ccc cca agt ctt ccc tcc tcc caa gcc agg ggt ggc aca ggg cac tag atc
3060
cct gga gtt cag gaa cca aca caa gca caa cca cgg gca taa gtt ggc ctt ggc cac tgc
3120
cac cca cgg ccc tcc ttt tgt gct cca tgc tgg cat ctt cac tcc cct acc cct tcc cca
3180
gcc act gct gct cat tca aac ttc tgt cca tgt ccc tcc act gtt cct atc agc agg tgg
3240
ccc ctg ggc atc aga aca gcc tgc cct ggg cac cag gtg gca gac aca ctc aga gca tgt
3300
ctg gct ttc ctg gtg ggt cca ggc tca ttc tgc ttc tga ttt ccc ctc ccc cag ggc tca
3360
ttt tcc ccc ttt ttc ctg tac aca tcc ctg tct acc tcc tct cac cct gcc aca gat tct
3420
tcc tat cac aca ggg atg cca gtt gta ttt gtg gg 3455

2005-05-05

FIGURE 3

Predicted Protein Sequence of Human ABCG4 Transporter

MAEKALEAVGCGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHITEAQR 50
 SHLPKRSAVDIEFVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGI 100
 MGPSGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTFRKMSCYIMQDD 150
 MLLPHLTVLEAMMVSANLNLTENPDVKNDLVTEILTALGLMSCSHRTAL 200
 LSGGQQRKRLAIALELVNPPVMEFDEPTSGLDSASCFQVVSMLKSLAQGG 250
 RTIICTIHQPSAKLFEMFDKLYILSQGQCIFKGVVTNLIPYLKGLGLHCP 300
 TYHNPADFIIIEVASGEYGDLPMLFRAVQNGLCAMAEKKSSPEKNEVPAP 350
 CPPCPPEVDPIESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMHV 400
 VIGVLIGLLYLHIGDDASKVFNNNTGCLFFSMLFLMFAALMPTVLTFFLEM 450
 AVFMREHLNYWYSLKAYYLAKTMADVFPQVVCVVYCSIVYWMTGQPAET 500
 SRFLFSALATATALVAQSLGLLIGAASNSLQVATFVGPVTAIPVLLFSG 550
 FFVSFKTIPTYLQWSSYLSYVRYGFEGVILTIYGMERGDLTCLERCPCR 600
 EPQSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLRVRVKSER 646

GPSTGAGKST LSGGQQRK VMEFDEPT Transmembrane domains are underline
 Walker A C signature Walker B

2004-05-20 10:00:00

FIGURE 4A

ClustalW Multiple Sequence Alignment of the Members of the ABCG Subfamily

```
ABCG1 1  MAAFSVGTAMNASSYSAEMTEPKSVCSVDEVSSNMEATETDLLNGHLKKVDNNLTEAQRFSSSLPRRAAVNIEFRD
77
ABCG4 1  MAEKALEAVGCGLGPGAVAMAVT-----LEDGAEPVLTTHLKKVENHITAQRFSHLPKRSAVDIEFVE
65
ABCG2 1  MSSSNVEVFIP-----VSQGTNGFPATVSN---DLKAFTEGAVLSFHNICYR- 45
ABCG5 1  MGDLSLTPGGSMGLQVNRG-----SQSSLEGAPATAPEP-----HSLGILHASYSVSHR-
50
ABCG8 1  MAGKAAEERGLPKGATPQDTSGLQDRLFS-----SESDNSLYFTYSGQPNTLEVR--DLNYQVDLASQVPWFELQALQ
70

ABCG1 78  LSYVPEGPWWRKKGYKTLKGISGKFNSGELVAIMGPSGAGKSTLMNLAGYRE-TG-MKGAVLINGLPRDLRCFRKVS
155
ABCG4 66  LSYVREGPCWRKRGYKTLKCLSGKFCRRELIGIMGPSGAGKSTFMNLAGYRE-SG-MKGQILVNGRPRELRTFRKMS
143
ABCG2 46  VKLKSGFLPCR-KPVEKEILSNINGIMKPG-LNAILGPTGGGKSSLLDVLAAARKDPSG-LSGDVLINGAPR-PANFKCNS
121
ABCG5 51  VRPWWITSCR-QQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCF
129
ABCG8 71  FKMPWTSPSCQ--NSCELGIONLSFKVRSGQMLAIIGSSGGRASLLDVITGRGHGGKIKSGQIWINQOPSSPQVLRKCV
148
      . . . . . : : : . . . . . : . . . :

ABCG1 156 CYIMQDDMLLPHLTVQEAMMVAHLKLQ--EKDEGRREMVKEILTALGLLSCANTRTGS-----LSGGQKRKLAIALELV
228
ABCG4 144 CYIMQDDMLLPHLTVLEAMMVSANLNT--ENPDVKNDLVTEILTALGLMSCSHRTAL-----LSGGQKRKLAIALELV
216
ABCG2 122 GYVQDDVVMGTLTVRENLOFSAAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELI
201
ABCG5 130 SYVLQSDTLSSSLTVRETLHYTALLAIRR-GNPGSFQKKVEAVMAELSLSHVADRLIGNYSLGGTSTGERRRVSIAAQLL
208
ABCG8 149 AHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLL
228
      . : . . . : : : . . . . . : : . . . : : . . . :

ABCG1 229 NNPPVMEFDEPTSGLDSASCFOVSLMKGLAQGGRSIICTIHQPSAKLFELFDQLYVLSQGCQVYRGVCNLPYLRDLG
308
ABCG4 217 NNPPVMEFDEPTSGLDSASCFOVSLMKSLAQGGRTIICTIHQPSAKLFEMFDKLYLSQGCQIFKGVVTNLIPYLKGLG
296
ABCG2 202 TDPSILSLDEPTTGLDSSSTANAVLLLLKRMSKQGRTIIFSIIHQPRYSIFKLFDLTLASGRLMFHGPQAELGYFESAG
281
ABCG5 209 QDPKVMLEFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCC
288
ABCG8 229 WNPGLILDEPTSGLDSFTAHLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLMTSGTPIYLGAAQHMVQYFTAIG
308
      : . . : : : : : . . . . . : : : : . . . : . . :

ABCG1 309 LNCPTYHNPADFVMEVASGEYGD--QN-SRLVRAVREGMCDSDHKRDLGGDAEVNPFVLRHPSEEVKQTKRLKGLRKDS-
384
ABCG4 297 LHCPTYHNPADFIIEVASGEYGD--LN-PMLFRAVQNGLCAMAEE-----KSSPEKNEVPAPCPPCPE-
357
ABCG2 282 YHCEAYNNPADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKP-LIEKLAEIYVNSSFYKETKAELHQLSGGEKKKK
360
```

20050505

FIGURE 4B

ABCG5 289 YPCPEHSNPFDFYMDLTSVDTQSKERE-IETSKRVQMIESAYKKS-----AICHKTLKNIERMKHLKTLP-
352

ABCG8 309 YPCPRYSNPADFYVDLTSIDRRSREQE-LATREKAQSLAALFLEKVR----DLDDFLWKAETKDLDEDTCESSVTPLD
382

: : : : .

ABCG1 385 SSMEGCHSFSASCLTQFCILFKRTFLSIMRDSVLTHLRITSHIGIGLLIGLLYLIGIGNETKK--VLSNSGFLFFSMLFLM
462

ABCG4 358 VDPIESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMHVVIGVLIGLLYLHIGDDASK--VFNN TGCLFFSMLFLM
435

ABCG2 361 ITVFKEISYTTSFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKNDSTG--IQNRAGVLFFLTNNOC
438

ABCG5 353 MVPFKTKD-SPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATP
431

ABCG8 383 TNCLPSPTKMPGAVQQFTTLIRQISNDFRDLPTLLIHGAEACLSMTIGFLYFGHGSIQLS--FMDTAALLFMIGALIP
460

: : .

ABCG1 463 FAALMPTVLTFFPLEMGVFLREHLNYWYSLKAYYLAKTMADVFPQIMFP-VAYCSIVYWMTSQPSDAVRVFLFAALGMTS
541

ABCG4 436 FAALMPTVLTFFPLEMAVFMREHLNYWYSLKAYYLAKTMADVFPQVVCV-VVYCSIVYWMTGQPAETSRFLFLSALATATA
514

ABCG2 439 FSSVS-AVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMRMLPSIIFTCIYVFMGLKPKADAFFVMMFTLMMVA
517

ABCG5 432 YTGMLNAVNLFPVLRVSDQESQDGLYQKWQMLLAYLHVLFPFSVVAT-MIFSSVCYWTGLHPEVARFGYFSAALLAPH
510

ABCG8 461 FNVILDVISKCYSERAMLYELEDGLYTGPFYFAKILGELPEHCAYI-IIYGMPTYWLANLRPGLQPFLLHFLLVVLV
539

: : : : : : :

ABCG1 542 LVAQSLGGLIG-AASTSLQVATFVGPTAIPVLLFSGFFVSFDTIPTYLQWMSYISYVRYGFEGVILSIYG-----L
612

ABCG4 515 LVAQSLGGLIG-AASNSLQVATFVGPTAIPVLLFSGFFVSFKTIPTYLQWSSYLSYVRYGFEGVILTIYG-----M
585

ABCG2 518 YSASSMALAIA-AGQSVVSVATLLMTICVFMMIFSGLLVNLTIIASWLSWLQYFSIPRYGFTALQHNEFLGQNFPCPLN
596

ABCG5 511 LIGEFLLVLLGIVQNPNIIVNSVALLSIAGVLVSGGFLRNIQEMPIPFKII SYFTFQKYCSEILVNEFYGLN---FTC
587

ABCG8 540 FCCRIMALAAA-ALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEGLMKIQFS-----R
610

: : : : : : :

ABCG1	613	DREDLHCDIDETCHFQ-KSEAILRELDVENAKLYLDFIVLGIFFI SLRLIAYLVLR YKIRAER	674
ABCG4	586	ERGD LTC-LEERC PFR-EPQSILRALDVEDAKLYMDFLVLGIFFLALRL LAYLVLR YRVKSER	646
ABCG2	597	ATGNNPC-NYATCTG--EEYLVKQGIDLS PWGLWKNHVALACMIVIFLT IAYLKLFLKKYS	655
ABCG5	588	GSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVVFKIRDHLISR	651
ABCG8	611	RTYKMPLGNLTIAVS---GDKILSVME LDSYPLYAIYLIVIGLSGGFMVLYYVSLRFIKQKPSQDW	673

FIGURE 5

ClustalW Multiple Sequence Alignment of Partial Human ABCG4 Transporter in GenBank
(AN: CAC17140) and Human ABCG4 Transporter of this Invention

ABCG4 vs. CAC17140

ABCG4	1	MAEKALEAVGCGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHITEAQRFSLPKRSAVDIEFVELSYSVREGPCWRKRG
80		
CAC17140	1	MAVTLEDGAEPVLTTHLKKVENHITEAQRFSLPKRSAVDIEFVELSYSVREGPCWRKRG
61		
	
ABCG4	81	YKTLKCLSGKFCRRELIGTMGSGAGKSTFMNILAGYRESGMKGQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLE
160		
CAC17140	62	YKTLKCLSGKFCRRELIGTMGSGAGKSTFMNILAGYRESGMKGQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLE
141		
	
ABCG4	161	AMMVSANLNLTENPDVKNDLVTEILTALGLMSCSHRTALSSGGQRKRLAIALELVNPPVMEFDEPTSGLDSASCFOVV
240		
CAC17140	142	AMMVSANLKLSEKQEVKKELVTEILTALGLMSCSHRTALSSGGQRKRLAIALELVNPPVMEFDEPTSGLDSASCFOVV
221		
	
ABCG4	241	SLMKSLAQGGRTIICTIHQPSAKLFEMFDKLYILSQGCIFKGVVTNLIPYLKGLGLHCPTYHNPADFIEVASGEYGD
320		
CAC17140	222	SLMKSLAQGGRTIICTIHQPSAKLFEMFDKLYILSQGCIFKGVVTNLIPYLKGLGLHCPTYHNPADFIEVASGEYGD
301		
	
ABCG4	321	NPMLFRAVQNGLCAMAEKKSSPEKNEVPAPCPPCPPEVDPIESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMSHV
400		
CAC17140	302	NPMLFRAVQNGLCAMAEKKSSPEKNEVPAPCPPCPPEVDPIESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMSHV
381		
	
ABCG4	401	VIGVLIGLLYLHIGDDASKVFNNNTGCLFFSMLFLMFAALMPTVLTFFPLEMAVFMREHLNYWYSLKAYYLAKTMADVFPQV
480		
CAC17140	382	VIGVLIGLLYLHIGDDASKVFNNNTGCLFFSMLFLMFAALMPTVLTFFPLEMAVFMREHLNYWYSLKAYYLAKTMADVFPQV
461		
	
ABCG4	481	VCPVVYCSIVYWMTGQPAETSRFLFSALATATALVAQSLGLLIGAASNSLQVATFVGPVTAIPVLLFSGFFVSFKTIPT
560		
CAC17140	462	VCPVVYCSIVYWMTGQPAETSRFLFSALATATALVAQSLGLLIGAASNSLQVATFVGPVTAIPVLLFSGFFVSFKTIPT
541		
	
ABCG4	561	YLOWSSYLSYVRYGFEGVILTIYGMERGLTCLEERCPFREPOSILRALDVEDAKLYMDFVLGIFFLALRLLAYLVRLY
640		
CAC17140	542	YLOWSSYLSYVRYGFEGVILTIYGMERGLTCLEERCPFREPOSILRALDVEDAKLYMDFVLGIFFLALRLLAYLVRLY
621		
	
ABCG4	641	RVKSER 646
CAC17140	622	RVKSER 627
	

Sequence Alignment

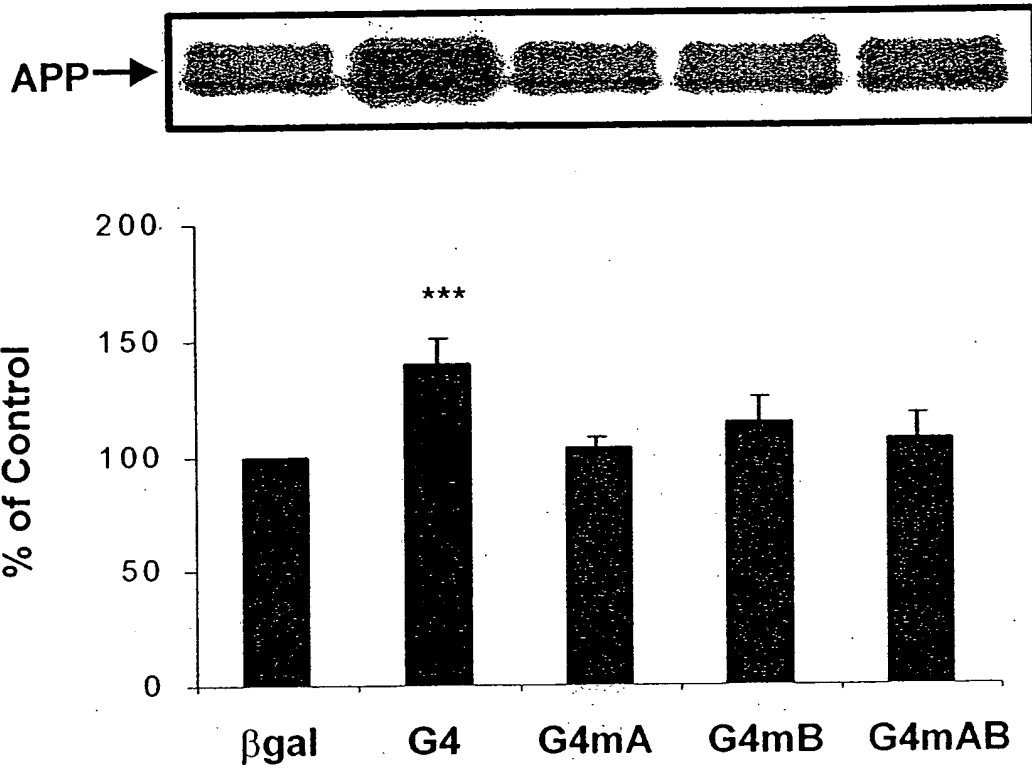


Figure 6

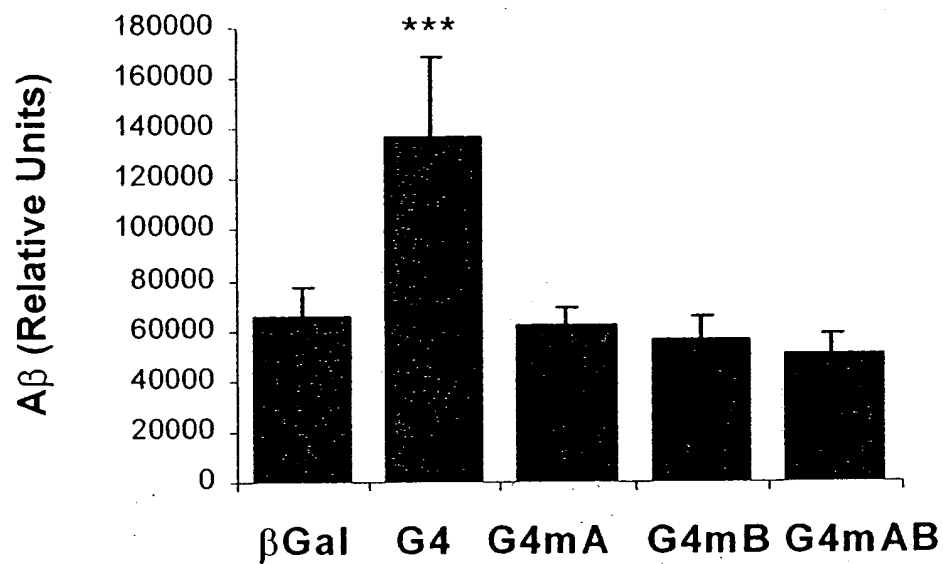


Figure 7

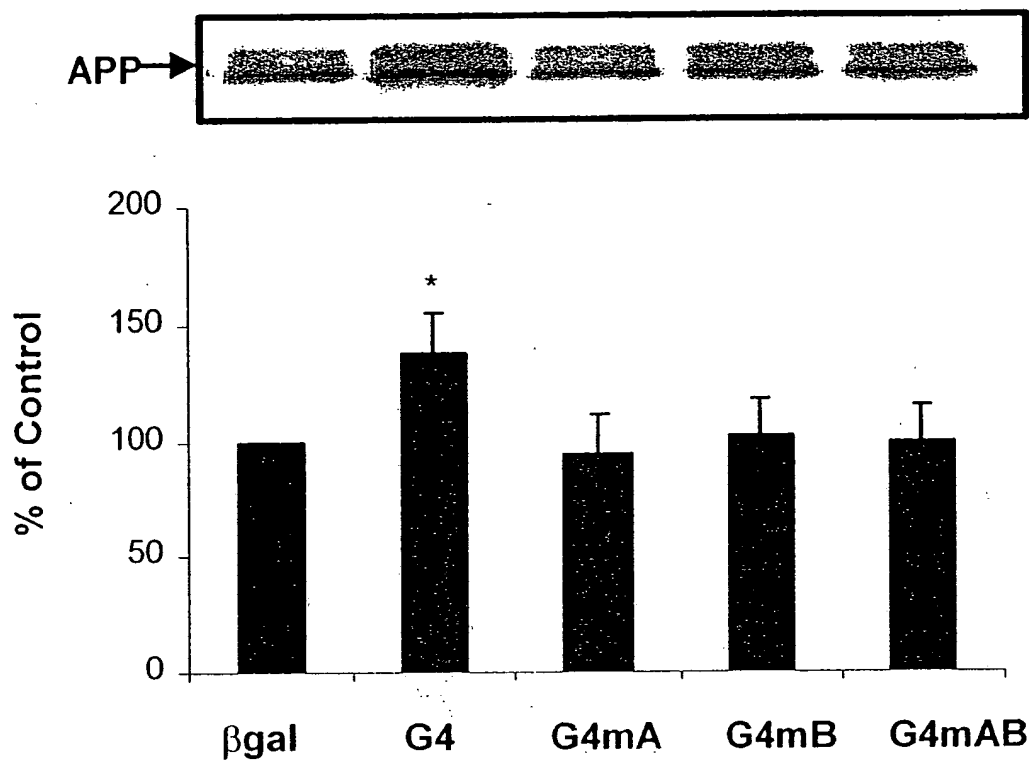


Figure 8

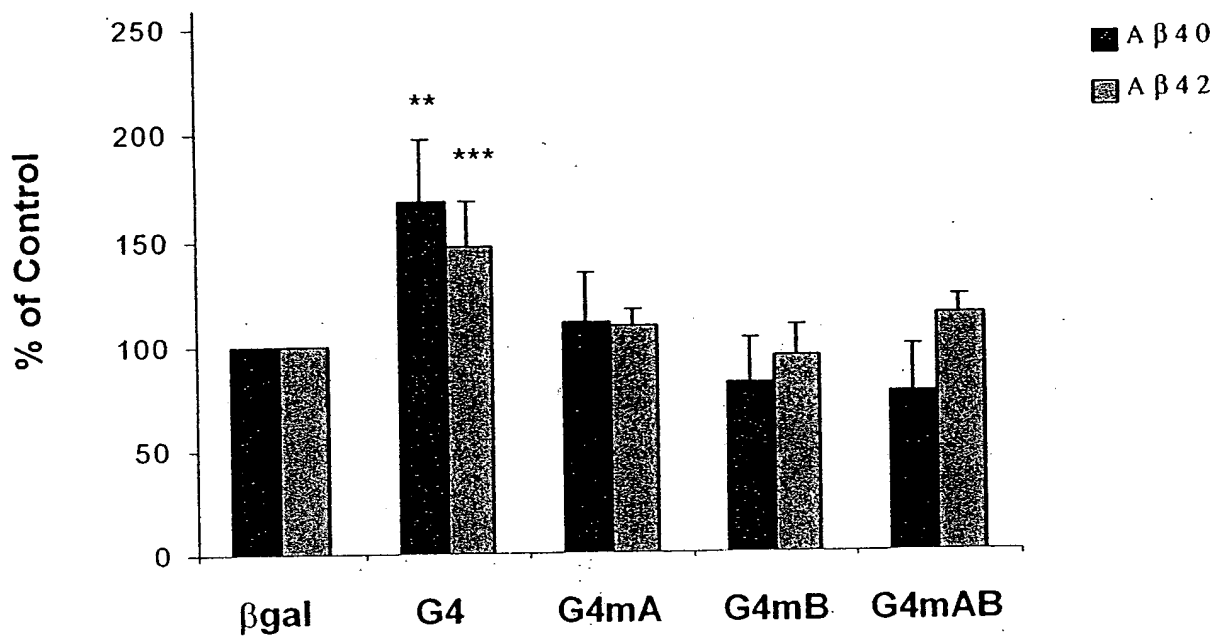


Figure 9

Directory: Mapleleaf/target clone/ABCG family/abcg4/g4cdna (Assembly file)

G4-clone nucleotide sequence range 1 to 2687

TACCGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTGCCCTTGCCACCATGCGGAGAAGGCGCTGGAGG
CCGTGGGCTGTGGACTAGGGCCGGGGGCTGTGGCCATGGCCGTGACGCTGGAGGACGGGGCGGAACCCCTG
TGCTGACCACGCACCTGAAGAAGGTGGAGAACCACATCACTGAAGCCCAGCGCTTCTCCCACCTACCCAAGC
GCTCAGCCGTGGACATCGAGTTCGTGGAGCTGTCTATTCGTGCGGGAGGGGCCCTGCTGGCGCAAAAGGG
GTTATAAGACCCTTCTCAAGTGCCTCTCAGGTAAATCTGCCGCCGGGAGCTGATTGGCATCATGGGCCCT
CAGGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGGAGTCTGGAATGAAGGGGCAGATCC
TGGTTAATGGAAGGCCACGGGAGCTGAGGACCTTCCGCAAGATGTCTGCTACATCATGAAGATGACATGC
TGCTGCCGCACCTCACGGTGTGGGAAGCCATGATGGTCTCTGTCTAACCTGAAGCTGAGTGAGAAGCAGGAGG
TGAAGAAGGAGCTGGTGACAGAGATCCTGACGGCACTGGGCCTGATGTCTGCTCCCGCACGAGGACAGCCC
TGCTCTCTGGCGGGCAGAGGAAGCGTCTGGCCATCGCCCTGGAGCTGGTCAACAACCCGCTGTCTGTTCT
TTGATGAGCCCACCAAGTGGTCTGGATAGCGCCTCTTGTTCCTCAAGTGGTGTCCCTCATGAAGTCCCTGGC
AGGGGGGCCGTACCATCATCTGCACCATCCACCAGCCAGTCCCAAGCTCTTTGAGATGTTTGACAAGCTCT
ACATCCTGAGCCAGGGTCAGTGCATCTTCAAAGGCGTGGTCACCAACCTGATCCCTATCTAAAGGGACTCG
GCTTGCAATTGCCCCACCTACCACAACCCGGCTGACTTCATCATCGAGGTGGCCTCTGGCGAGTATGGAGACC
TGAACCCCATGTTGTTTCAGGGCTGTGCAGAATGGGCTGTGCGCTATGGCTGAGAAGAGAGCAGCCCTGAGA
AGAACGAGGTCCCTGCCCCATGCCCTCCTTGTCTCCGGAAGTGGATCCCATTTGAAAGCCACACCTTTGCCA
CCAGCACCTCACACAGTTCTGCATCCTCTTCAAGAGGACCTTCCTGTCCATCCTCAGGGACACGGTCTCTGA
CCCACCTACGGTTTCATGTCCCACGTGGTTATTGGCGTGCTCATCGGCCTCCTCTACCTGCATATTGGCGACG
ATGCCAGCAAGGTCTTCAACAACACCGGCTGCCTCTTCTCTCCATGCTGTTCCCTCATGTTCCCGCCCTCA
TGCCAACTGTGCTCACCTTCCCCTTAGAGATGGCGGTCTTCATGAGGGAGCACCTCAACTACTGGTACAGCC
TCAAAGCGTATTACCTGGCCAAGACCATGGCTGACGTGCCCTTTCAGGTGGTGTGTCCGGTGGTCTACTGCA
GCATTGTGTACTGGATGACGGGCCAGCCCGCTGAGACCAGCCGCTTCCTGCTCTTCTCAGCCCTGGCCACCG
CCACCGCCTTGGTGGCCCAATCTTTGGGGCTGCTGATCGGAGCTGCTTCCAACCTCCCTACAGGTGGCCACTT
TTGTGGGCCCAGTTACCGCCATCCCTGTCTCTTGTCTCCGGCTTCTTTGTCTAGCTTCAAGACCATCCCCA
CTTACCTGCAATGGAGCTCCTATCTCTCTATGTGAGGTATGGCTTTGAGGGTGTGRTCTGACGATCTATG
GCATGGAGCGAGGAGACCTGACATGTTTAGAGGAACGCTGCCMGTTCGGGAGCCACAGAGCATCCTCCGAG
CGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTCCTGGTCTTGGGCATCTTCTTCTAGCCCTGCGGC
TGCTGGCCTACCTTGTGCTGCGTTACCGGGTCAAGTCAGAGAGATAGAGGCTTGCCCCAGCCTGTACCCAG
CCCCTGCAGCAGGAAGCCCCCAGTCCCAGCCCTTTGGGACTGTTTTAACCTTATAGACTTGGGCACTGGTTC
CTGGCGGGGCTATCCTCTCCTCCCTTGGCTCCTCCACAGGCTGGCTGTGCGACTGCGCTCCAGCCTGGGCT
CTGGGAGTGGGGGCTCCAGCCCTCCCCACTATGCCCAGGAGTCTTCCCAAGTTGATGCGGTTTGTAGCTTCC
TCCCTACTCTCTCCAACACCTGCATGCAAAGACTACTGGGAGGCTGCTGCCTCCTTCCCTGCCCATGGCACC
TCCTCTGCTGTCTGCCTGGGAGCCCTAGGCTCTCTAGGGCCCCACTTACAACCTGACCAAAGTGGCCCCCTCT
KGGGGTCCCCACCACACAAGTGTGTTGTAAGTGGGCTGCTATAAGGTTGGAGTTCAGGGCTGGGCCCCTGGT
GGAGTCCACTGGAAGTCCCATCATGGATGTTGAAATGGACAGGGAAGGACTCTGGAAGTCTCTTCTCTCTCC
TCCTCTTCTCTCCACCCCTAGACCCTGGCTGACTTGGACAATCTGCCAGGACAGAAGCTGGGGTTTTCTGTCT
TAGGTCAACCACTCCCAATCCTGGGGGRTTGGAGRGGCCTGGGGSTGTGGGRTGSCCATCCCCCTCCCCATC
ACCTTGGTGGGGGSAGGGCCTG

Figure 10

G4-clone polypeptid sequence range 1 to 646

MAEKALEAVGCGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHITEAQRFSHLPKRSAVD
IEFVELSYSVREGPCWRKRGYKTLLKCLSGKFCRRELIGIMGPSGAGKSTFMNILAGYRE
SGMKGQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLEAMMVSA NLK LSEKQEVKKEL
VTEILTALGLMSCSRTRTALLSGGQRKRLAIALELVNNPPVMFFDEPTSGLDSASCFQVV
SLMKSLAQGGRTI ICTI HQPSAKLFEMFDKLYILSQGQCI FKG VVTNLI PYLKGLGLHCP
TYHNPADFI IEVASGEYGD LN PMLFRAVQNGLCAMAEKKSSPEKNEVPAPCPPCPPEVDP
IESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFM SHV VIGVLIGLLYLHIGDDASKV
FNNTGCLFFSMLFLMFAALMPTVLTFPLEMAVFMREHLNYWYSLKAYYLAKTMADV PFQV
VCPV VYCSIVYWMTGQPAETS RFL LFSALATATALVAQSLG LLIGAASNSLQVATFVGPV
TAIPVLLFSGFFVSFKTIPTYLQWSSYLSYVRYGFEGVXLT IYGMERGD LTCLEERCXFR
EPQSILRALDVEDAKLYMDFLVLGIFFLALRL LAYLVLRYRVKSER

2004-05-25 10:04:55

Figure 11